SemRevRec: A Recommender System based on User Reviews and Linked Data

Iacopo Vagliano[a], Diego Monti[b] and Maurizio Morisio[b]

[a]ZBW - Leibniz Information Centre for Economics, Kiel, Germany
[b]Dipartimento di Automatica e Informatica, Politecnico di Torino, Turin, Italy

Introduction

Currently, most of recommender systems exploit user ratings to infer preferences, although the growing popularity of social and e-commerce websites has encouraged users to write textual reviews.
These reviews enable recommender systems to represent the multi-faceted nature of users' opinions and build a fine-grained preference model, which cannot be obtained from overall ratings [1].
We argue that the information extracted from user reviews, combined with Linked Data, can be exploited in recommendation tasks.

- Linked Data can provide a semantically rich representation of the items to recommend.
- Reviews may be a source of not obvious connections that are hardly available in a KB.

For instance, various reviews of *Interstellar* mention Stanley Kubrick, although there is not a direct link between these two resources in DBpedia.
We propose a novel recommendation approach based on the semantic annotation of reviews to extract useful information from them. A preliminary offline study suggests that our method provides better prediction and ranking accuracy than another recommender system based on Linked Data, while it increases the diversity of recommendations with respect to all the techniques considered.

Architecture

The architecture of the whole system is depicted in Figure 1. It consists of two main modules which are highlighted with different colors.

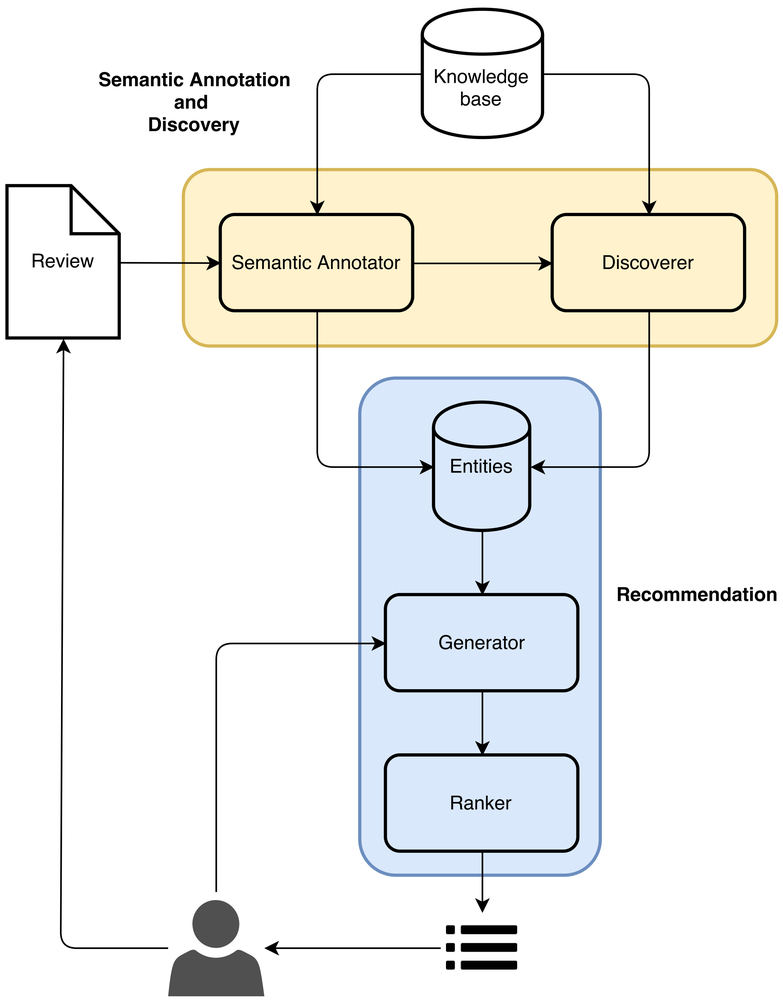

Figure 1: Architecture of the system

Semantic Annotation and Discovery

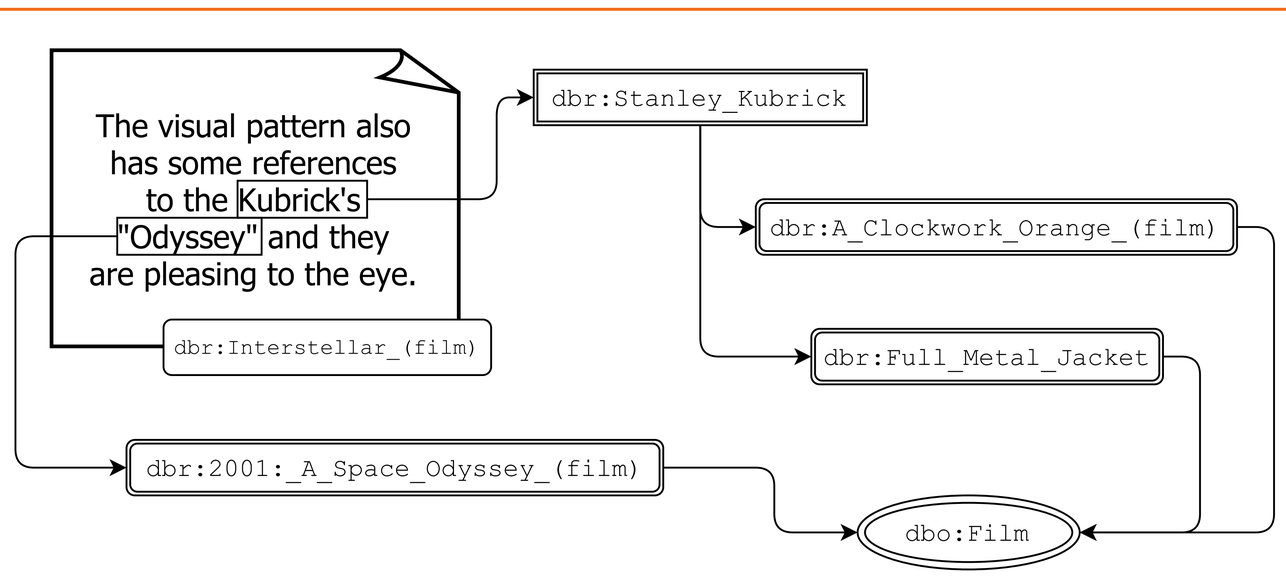

Figure 2: Annotation of a review about *Interstellar*

Recommendation

The recommendation process is based on the generation of the candidate items and their ranking:

1. the system selects the annotated entities which were mentioned in the reviews of the initial item;
2. it obtains the entities which mention the initial item;
3. it retrieves the entities discovered through the initial item and vice versa;
4. it ranks the candidate recommendations according to the following equation.

$$R(i) = \beta \cdot \frac{\alpha \cdot occurrence_i}{max_{j \in CR}(occurrence_j)} + \gamma \cdot (1 - LDSD(i, i_o))$$

Evaluation

We evaluated SemRevRec with a preliminary offline experiment conducted in the movie domain. We annotated the reviews available on IMDb for the top-250 movies. We relied on the MovieLens 1M dataset for obtaining the actual user ratings.
The evaluation was performed with LibRec. We executed a 5-fold cross-validation considering as positive the ratings greater than 3 on a scale from 1 to 5. Using the top-10 recommendations for each user, we computed the measures of precision, recall, nDCG, Entropy Based Novelty (EBN) [2], and diversity [3]. We compared our technique with the Most Popular and the Random Guess baseline algorithms, and with SPrank [4]. We configured SPrank to exploit LambdaMart as ranking method and the properties related to the movie domain (`dct:subject`, `dbo:director`, and `dbo:starring`).
Table 1 lists the results of the experiment. For all the measures but EBN, higher values mean better results, while the lower is EBN, the higher is the novelty.

Table 1: Results of the experiment

Algorithm	Precis.	Recall	nDCG	EBN	Divers.
SemRevRec	0.0882	0.0459	0.0589	1.5671	0.1838
SPrank	0.0584	0.0327	0.0409	0.8244	0.1551
Popular	0.1325	0.0840	0.0969	2.7439	0.1412
Random	0.0055	0.0028	0.0031	0.3018	0.1679

Future Work

As future work, we plan to:

- investigate if increasing the number of reviews improves the quality of the recommendations;
- consider the sentiment and the linking confidence associated with the annotated entities.

References

[1] Li Chen, Guanliang Chen, and Feng Wang.
Recommender systems based on user reviews: The state of the art.
User Modeling and User-Adapted Interaction, 25(2):99–154, 2015.

[2] Alejandro Bellogìn, Ivàn Cantador, and Pablo Castells.
A study of heterogeneity in recommendations for a social music service.
In *Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, HetRec '10, pages 1–8. ACM.

[3] Mi Zhang and Neil Hurley.
Avoiding monotony: Improving the diversity of recommendation lists.
In *Proceedings of the 2008 ACM Conference on Recommender Systems*, RecSys '08, pages 123–130, New York, NY, USA, 2008. ACM.

[4] Tommaso Di Noia, Vito Claudio Ostuni, Paolo Tomeo, and Eugenio Di Sciascio.
SPrank: Semantic path-based ranking for top-n recommendations using linked open data.
ACM Transactions on Intelligent Systems and Technology, 8(1):9:1–9:34, 2016.